SUT Journal of Mathematics
Vol. 45, No. 2 (2009), 119–135

Asymptotic approximation of EPMC for linear discriminant analysis using ridge type estimator in high-dimensional data with fewer observations

Masashi Hyodo

(Received August 6, 2009; Revised December 10, 2009)

Abstract. In this paper, the problem of classifying a new observation vector into one of the two normal populations for high-dimensional data is considered. High-dimensional data means that the total number of observation vectors from the two groups is less than the dimension of the observation vectors. Recently, linear discriminant analysis (LDA) for high-dimensional data such as microarray data has been considered. A simple way is to use the Moore-Penrose inverse when the sample covariance matrix is singular. In this paper, we suggest another type LDA approach for high-dimensional data. This method is based on a ridge type estimator of covariance matrix which was proposed by Srivastava and Kubokawa (2008). In addition, we derive asymptotic approximation of EPMC for this method in the situation of $n = O(p^{\delta})$, $p \to \infty$, $0 < \delta < 1/2$.

AMS 2000 *Mathematics Subject Classification.* 62H12, 62E30.

Key words and phrases. asymptotic approximations, expected probability of misclassification, high dimensional data, linear discriminant function, ridge estimator.

§1. Introduction

We deal with the problem of classifying a $p \times 1$ observation vector $\boldsymbol{x}$ as coming from one of two populations Π_1 and Π_2. Let Π_i, $i = 1, 2$ have p-variate normal populations with mean vector $\boldsymbol{\mu}_i$ and the common positive definite covariance matrix Σ, where $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$. Assume that random sample vectors $\boldsymbol{x}_{ij}, j = 1, \ldots, N_i$ from Π_i, $i = 1, 2$ are given. Consider the case in which all parameters are unknown. linear discriminant analysis (LDA) is one of the standard classical methods for classifying $\boldsymbol{x}$ into either Π_1 or Π_2, which is given as follows:

$$W = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' S^{-1} \{\boldsymbol{x} - \tfrac{1}{2}(\bar{\boldsymbol{x}}_1 + \bar{\boldsymbol{x}}_2)\} \gtrless 0 \Longrightarrow \quad \boldsymbol{x} \in \Pi_1(\Pi_2).$$

Here, $\bar{\boldsymbol{x}}_1$, $\bar{\boldsymbol{x}}_2$ and S are the sample mean vectors and the pooled sample covariance matrix given by

$$\bar{\boldsymbol{x}}_i = N_i^{-1} \sum_{j=1}^{N_i} \boldsymbol{x}_{ij}, \quad i = 1, 2,$$

$$S = n^{-1} \sum_{i=1}^{2} \sum_{j=1}^{N_i} (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)',$$

respectively, where $n = N_1 + N_2 - 2$. It is generally difficult to obtain an explicit expression for the expected probabilities of misclassification (EPMC), that is, the probabilities of misclassifying $\boldsymbol{x}$ into Π_2 (Π_1) when it actually belongs to Π_1 (Π_2). So, there are much works for their asymptotic approximations. Type-I approximations are the ones under a framework such that N_1 and N_2 are large and p is fixed. For a review of these results, see, e.g., Siotani (1982). Further, the ones under a framework that N_1, N_2 and p are all large have also been studied (see, e.g., Raudys (1972), Fujikoshi and Seo (1998)). Moreover, Fujikoshi (2000) gave explicit formula of error bounds for approximation of EPMC proposed by Lachenbruch (1968).

Recently, linear discriminant analysis for high-dimensional data has been considered. A simple way is to use the Moore-Penrose inverse when the sample covariance matrix is singular. On the other hand, the usefulness of the ridge type estimators has been recognized by Srivastava and Kubokawa (2007). In order to guarantee the nonsingularity of S, we use the following ridge type estimator instead of S.

$$S_r = S + \lambda I.$$

From Srivastava and Kubokawa (2007) and Kubokawa and Srivastava (2008), the following ridge parameter is chosen by the empirical Bayes method:

$$\lambda = \frac{\sqrt{p}\hat{a}_1}{n}, \; \hat{a}_1 = \frac{\mathrm{tr}(S)}{p}.$$

Using above estimator, we suggest ridge type linear discriminant analysis (RTLDA);

$$W_r = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' S_r^{-1} \{\boldsymbol{x} - \tfrac{1}{2}(\bar{\boldsymbol{x}}_1 + \bar{\boldsymbol{x}}_2)\} \lesseqgtr 0 \quad \Longrightarrow \quad \boldsymbol{x} \in \Pi_1(\Pi_2). \tag{1.1}$$

In this paper, we consider an asymptotic approximation of the EPMC for large p with $n = O(p^{\delta})$, $0 < \delta < 1/2$. The EPMC for the RTLDA may be expressed as follows:

$$e(2|1) = \Pr(W_r \leq 0 | \boldsymbol{x} \in \Pi_1), \quad e(1|2) = \Pr(W_r \geq 0 | \boldsymbol{x} \in \Pi_2).$$

The organization of this paper is as follows. In Section 2, we give an asymptotic approximation of EPMC for RTLDA and derive an estimator of EPMC. Further we evaluate our results in Section 2 numerically by Monte Carlo simulations in Section 3. In Section 4, we investigate EPMC of RTLDA for Leukemia dataset which were considered by Dudoit et al. (2002). The conclusion of our study is summarized in Section 5.

§2. Asymptotic approximation of EPMC for RTLDA

In this section, we consider an asymptotic approximation for RTLDA under the following assumptions:

$$\text{A1}: \quad n = O(p^{\delta}),\ N_i = O(p^{\delta}),\ p \to \infty,\ 0 < \delta < 1/2,\ i = 1, 2.$$

Further, in addition to A1, we assume the following assumptions:

$$\begin{aligned} \text{A2}: &\quad \operatorname{tr} \Sigma^i/p \to a_{i0},\ 0 < a_{i0} < \infty,\ i = 1, \dots, 6, \\ \text{A3}: &\quad 0 < \boldsymbol{\delta}'\boldsymbol{\delta}/p < \infty,\ \boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, \\ \text{A4}: &\quad 0 < \boldsymbol{\delta}'\Sigma\boldsymbol{\delta}/p < \infty. \end{aligned}$$

The EPMC based on the rule (1.1) are expressed as

$$e(2|1) = \Pr(W_r < 0 | \boldsymbol{x} \in \Pi_1), \quad e(1|2) = \Pr(W_r > 0 | \boldsymbol{x} \in \Pi_2).$$

Since $e(1|2)$ is given from $e(2|1)$ by interchanging N_1 and N_2, we only deal with $e(2|1)$. Let the statistics V, Z, U be defined as follows (see e.g., Fujikoshi (2000)):

$$\begin{aligned} V &= (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' S_r^{-1} \Sigma S_r^{-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2), \\ Z &= V^{-\frac{1}{2}} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' S_r^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_1), \\ U &= (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' S_r^{-1} (\bar{\boldsymbol{x}}_1 - \boldsymbol{\mu}_1) - \tfrac{1}{2} D^2. \end{aligned}$$

Here $D^2 = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' S_r^{-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)$. Then, it may be expressed that

$$W_r = V^{-1/2} Z - U$$

under $\boldsymbol{x} \in \Pi_1$. Since Z and (U, V) are independent, and Z is distributed according to $N(0, 1)$ (here after, denoted by $Z \sim N(0, 1)$),

$$e(2|1) = \mathrm{E}_{(U,V)}[\Phi(U/\sqrt{V})],$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of $N(0, 1)$. To evaluate the expectation with respect to U and V explicitly, set

$$\begin{aligned} \boldsymbol{z}_1 &= N^{-\frac{1}{2}} (N_1 \bar{\boldsymbol{x}}_1 + N_2 \bar{\boldsymbol{x}}_2 - N_1 \boldsymbol{\mu}_1 - N_2 \boldsymbol{\mu}_2), \\ \boldsymbol{z}_2 &= \left(\frac{N}{N_1 N_2} \right)^{-\frac{1}{2}} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2 - \boldsymbol{\mu}_1 + \boldsymbol{\mu}_2), \end{aligned}$$

where $N = n + 2$. Note that $\boldsymbol{z}_i \sim N_p(\boldsymbol{0}, \Sigma)$, $i = 1, 2$. In addition, $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ are independent. We can express U and V in terms of $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ as the following:

$$
\begin{aligned}
U = & -\frac{1}{2}\boldsymbol{\delta}' S_r^{-1}\boldsymbol{\delta} + \frac{1}{N^{\frac{1}{2}}}\boldsymbol{\delta}' S_r^{-1}\boldsymbol{z}_1 - \left(\frac{N_1}{NN_2}\right)^{\frac{1}{2}} \boldsymbol{\delta}' S_r^{-1}\boldsymbol{z}_2 \\
& + \frac{1}{(N_1N_2)^{\frac{1}{2}}}\boldsymbol{z}_1' S_r^{-1}\boldsymbol{z}_2 - \frac{N_1 - N_2}{2N_1N_2}\boldsymbol{z}_2' S_r^{-1}\boldsymbol{z}_2, \\
V = & \, \boldsymbol{\delta}' S_r^{-1}\Sigma S_r^{-1}\boldsymbol{\delta} + 2\left(\frac{N}{N_1N_2}\right)^{\frac{1}{2}} \boldsymbol{\delta}' S_r^{-1}\Sigma S_r^{-1}\boldsymbol{z}_2 + \frac{N}{N_1N_2}\boldsymbol{z}_2' S_r^{-1}\Sigma S_r^{-1}\boldsymbol{z}_2.
\end{aligned}
$$

We propose an approximation of EPMC for RTLDA as follows:

$$e(2|1) \approx \Phi(\xi), \tag{2.1}$$

where $\xi \in \mathcal{R}$ s.t. $|\Phi(U/\sqrt{V}) - \Phi(\xi)| = o_p(1)$. Here, the notation $o_p(p^i)$ denotes a term less than the i-th order with respect to p^i. To find ξ, we use the following lemmas.

Lemma 1 (Srivastava (2005))**.** *Let $nS \sim W_p(\Sigma, n)$. Then,*

(i) $\mathrm{E}[\hat{a}_i] = a_i$ *for* $i = 1, 2$.

(ii) $\lim_{p\to\infty} \hat{a}_i = a_{i0}$ *in probability for* $i = 1, 2$.

(iii) $\mathrm{Var}(\hat{a}_1) = 2a_2/(pn)$.

Here, $\hat{a}_1 = \mathrm{tr}(S)/p$, $\hat{a}_2 = n^2/\{(n-1)(n+2)\}\{\mathrm{tr}(S^2)/p - (\mathrm{tr}(S))^2/(np)\}$.

Lemma 2 (Srivastava (2007))**.** *Let $nS \sim W_p(\Sigma, n)$, $n < p$, and $nS = H_1' L H_1$, where $H_1' H_1 = I_n$ and $L = (\ell_1, \ldots, \ell_n)$, an $n \times n$ diagonal matrix which contains the non-zero eigenvalues of V. Then,*

(i) $\lim_{p\to\infty} \frac{L}{p} = a_{10} I_n$ *in probability.*

(ii) $\lim_{p\to\infty} H_1' \Sigma H_1 = \frac{a_{20}}{a_{10}} I_n$ *in probability.*

(iii) $\lim_{p\to\infty} H_1' \Sigma^2 H_1 = \frac{a_{30}}{a_{10}} I_n$ *in probability.*

(iv) $\lim_{p\to\infty} \frac{\boldsymbol{a}' H_1 H_1' \boldsymbol{a}}{n} = \frac{\boldsymbol{a}' \Sigma \boldsymbol{a}}{p}$ *in probability for* $\boldsymbol{a} \in \mathcal{R}^p$.

(v) $\lim_{p\to\infty} \frac{\boldsymbol{a}' H_1 H_1' \Sigma \boldsymbol{a}}{n} = \frac{\boldsymbol{a}' \Sigma^2 \boldsymbol{a}}{p}$ *in probability for* $\boldsymbol{a} \in \mathcal{R}^p$.

For the proofs of Lemma 1 and Lemma 2 except (iii) and (v), see Srivastava (2005, 2007). About (iii) and (v), we can easily show it by using the method how is similar to proofs of (ii) and (iv) in Lemma 2. Using Lemmas 1 and 2, following lemma is derived.

Lemma 3. *Under the assumption* A1-A4*, it holds that*

(i) $U/p^{\delta+1/2} = -\dfrac{n}{2p^{\delta}}\left(\dfrac{\boldsymbol{\delta}'\boldsymbol{\delta}}{pa_{10}} + \dfrac{N_1 - N_2}{N_1 N_2}\right) + o_p(p^{-1/2})$.

(ii) $V/p^{2\delta} = \dfrac{n^2}{p^{2\delta}}\left(\dfrac{\boldsymbol{\delta}'\Sigma\boldsymbol{\delta}}{pa_{10}^2} + \dfrac{Na_{20}}{N_1 N_2 a_{10}^2}\right) + o_p(p^{-1/2})$.

The proof of Lemma 3 stated are given in Appendix. From Lemma 3, we can get

$$\left|\frac{U}{\sqrt{V}} - \xi\right| = o_p(1), \tag{2.2}$$

where

$$\begin{aligned}
&\xi = -\frac{\sqrt{p}u_0}{2\sqrt{v_0}},\\
&u_0 = \frac{\Delta_0}{a_{10}} + \frac{N_1 - N_2}{N_1 N_2},\ v_0 = \frac{\Delta_1}{a_{10}^2} + \frac{Na_{20}}{N_1 N_2 a_{10}^2},\\
&\Delta_0 = \frac{\boldsymbol{\delta}'\boldsymbol{\delta}}{p},\ \Delta_1 = \frac{\boldsymbol{\delta}'\Sigma\boldsymbol{\delta}}{p}.
\end{aligned}$$

On the other hand, it is noted that

$$\begin{aligned}
&|\Phi(U/\sqrt{V}) - \Phi(\xi)|\\
&= \int_{\min(U/\sqrt{V},\xi)}^{\max(U/\sqrt{V},\xi)} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx\\
&\le |\max(U/\sqrt{V},\xi) - \min(U/\sqrt{V},\xi)| \times \frac{1}{\sqrt{2\pi}} e^{-\frac{\{\max(U/\sqrt{V},\xi)\}^2}{2}}\\
&\le |U/\sqrt{V} - \xi| \times \frac{1}{\sqrt{2\pi}}.
\end{aligned}$$

From (2.2), we get following theorem.

Theorem 1. *Under the assumption* A1-A4*, it holds that*

$$\lim_{p\to\infty} |\Phi(U/\sqrt{V}) - \Phi(\xi)| = 0 \textit{ in probability.}$$

Further, we consider $|e(2|1) - \Phi(\xi)|$. It can be expressed as

$$\begin{aligned} |e(2|1) - \Phi(\xi)| &= |\,\mathrm{E}[\Phi(U/\sqrt{V})] - \Phi(\xi)| \\ &= |\,\mathrm{E}[\Phi(U/\sqrt{V}) - \Phi(\xi)]| \\ &\leq \mathrm{E}[|\Phi(U/\sqrt{V}) - \Phi(\xi)|]. \end{aligned}$$

From $0 < \mathrm{E}[|\Phi(U/\sqrt{V}) - \Phi(\xi)|^2] < \infty$ and Theorem 1,

$$\lim_{p\to\infty} \sup_{\Theta} \mathrm{E}[|\Phi(U/\sqrt{V}) - \Phi(\xi)|] = \mathrm{E}[\lim_{p\to\infty} \sup_{\Theta} |\Phi(U/\sqrt{V}) - \Phi(\xi)|] = 0,$$

where $\Theta = \{\boldsymbol{\mu}_1,\ \boldsymbol{\mu}_2,\ \Sigma | 0 < a_{i0} < \infty,\ i = 1, \ldots, 6,\ 0 < \boldsymbol{\delta}'\boldsymbol{\delta}/p < \infty,\ 0 < \boldsymbol{\delta}'\Sigma\boldsymbol{\delta}/p < \infty\}$. Thus, we can get

$$\lim_{p\to\infty} \sup_{\Theta} |e(2|1) - \Phi(\xi)| = 0.$$

So, we suggest an approximation of $e(2|1)$ as follows:

$$e(2|1) \approx \Phi\left(\xi\right). \tag{2.3}$$

Next, we consider an estimator of $e(2|1)$. u_0 and v_0 include the unknown parameters $a_{i0},\ \Delta_{i-1}$ for $i = 1, 2$, which are estimated by the consistent estimators

$$\begin{aligned} \hat{a}_{10} &= \frac{\mathrm{tr}(S)}{p},\ \hat{a}_{20} = \frac{n^2}{(n-1)(n+2)} \left\{ \frac{\mathrm{tr}(S^2)}{p} - \frac{(\mathrm{tr}(S))^2}{np} \right\}, \\ \hat{\Delta}_0 &= \frac{(\boldsymbol{x}_1 - \boldsymbol{x}_2)'(\boldsymbol{x}_1 - \boldsymbol{x}_2)}{p} - \frac{N_1 + N_2}{N_1 N_2} \hat{a}_{10}, \\ \hat{\Delta}_1 &= \frac{(\boldsymbol{x}_1 - \boldsymbol{x}_2)' S (\boldsymbol{x}_1 - \boldsymbol{x}_2)}{p} - \frac{N_1 + N_2}{N_1 N_2} \hat{a}_{20}. \end{aligned}$$

Replacing the unknown values with their consistent estimator, we can propose an estimator of $e(2|1)$, which is given in the following result:

$$\hat{e}(2|1) = \Phi(\hat{\xi}), \tag{2.4}$$

where

$$\hat{\xi} = \frac{\sqrt{p}\hat{u}_0}{2\sqrt{\hat{v}_0}},\ \hat{u}_0 = \frac{\hat{\Delta}_0}{\hat{a}_{10}} + \frac{N_1 - N_2}{N_1 N_2},\ \hat{v}_0 = \frac{\hat{\Delta}_1}{\hat{a}_{10}^2} + \frac{N \hat{a}_{20}}{N_1 N_2 \hat{a}_{10}^2}.$$

§3. Simulation Studies

We are interested in the accuracy of the asymptotic approximations for EPMC proposed in (2.3) and estimator for EPMC given in (2.4). We generate the datasets as follows:

$$\begin{aligned}\Pi_1 &: \boldsymbol{x}_{11}, \boldsymbol{x}_{12}, \ldots, \boldsymbol{x}_{1N_1} \overset{i.i.d.}{\sim} N_p(\boldsymbol{\mu}_1, \Sigma),\\ \Pi_2 &: \boldsymbol{x}_{21}, \boldsymbol{x}_{22}, \ldots, \boldsymbol{x}_{2N_2} \overset{i.i.d.}{\sim} N_p(\boldsymbol{\mu}_2, \Sigma),\end{aligned}$$

where

$$\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_p)\mathcal{R}\,\mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_p);\ \mathcal{R} = \left(\rho^{|i-j|}\right)$$

for $\rho = 0.1,\ 0.4$ or 0.8 and $\sigma_i = 2 + (p - i + 1)/p$. Note that the assumption A2 does not hold for the case $\rho = 0.8$. The mean vector of the first group was chosen as

$$\boldsymbol{\mu}_1 = (\mu_1, \mu_2, \ldots, \mu_p)',\ \mu_i = (-1)^i(c + u_i), i = 1, \ldots, p$$

for random variable u_i from a uniform distribution on the interval $[0, 1]$ and $c = 0.2$ or 0.5. We chose the p dimensional mean vector of the second group as a zero vector, i.e. $\boldsymbol{\mu}_2 = (0, 0, \ldots, 0)'$. We report the results corresponding to: $(N_1, N_2) = (10, 10),\ (15, 5),\ (5, 15)$ when $p = 100$ or 200. Besides, the true values of EPMC in tables are average values of 10,000 repetitions. We consider the following two values:

$$\begin{aligned}\text{Approx}: &\ \Phi(\xi),\\ \text{Est}: &\ \mathrm{E}[\Phi(\hat{\xi})].\end{aligned}$$

We examine the effectiveness of this approximation by checking how close Approx and Est are to the true value.

Table 1. The accuracy of Approx and Est ($c = 0.2$)

(p, N_1, N_2)	ρ	True value	Approx	Est
(100,10,10)	0.1	0.221	0.207	0.240
	0.4	0.210	0.225	0.252
	0.8	0.179	0.323	0.381
(100,15,5)	0.1	0.041	0.029	0.054
	0.4	0.053	0.042	0.076
	0.8	0.084	0.171	0.264
(100,5,15)	0.1	0.634	0.644	0.678
	0.4	0.561	0.611	0.619
	0.8	0.437	0.582	0.561

Table 2. The accuracy of Approx and Est ($c = 0.5$)

(p, N_1, N_2)	ρ	True value	Approx	Est
(100,10,10)	0.1	0.090	0.075	0.098
	0.4	0.079	0.069	0.117
	0.8	0.017	0.087	0.153
(100,15,5)	0.1	0.019	0.016	0.025
	0.4	0.013	0.012	0.024
	0.8	0.014	0.077	0.172
(100,5,15)	0.1	0.364	0.372	0.404
	0.4	0.327	0.383	0.411
	0.8	0.192	0.435	0.461

Table 3. The accuracy of Approx and Est ($c = 0.2$)

(p, N_1, N_2)	ρ	True value	Approx	Est
(200,10,10)	0.1	0.130	0.124	0.174
	0.4	0.127	0.112	0.181
	0.8	0.149	0.240	0.354
(200,15,5)	0.1	0.006	0.004	0.021
	0.4	0.007	0.007	0.024
	0.8	0.048	0.081	0.225
(200,5,15)	0.1	0.673	0.696	0.678
	0.4	0.610	0.645	0.621
	0.8	0.516	0.616	0.571

Table 4. The accuracy of Approx and Est ($c = 0.5$)

(p, N_1, N_2)	ρ	True value	Approx	Est
(200,10,10)	0.1	0.033	0.027	0.055
	0.4	0.019	0.021	0.045
	0.8	0.035	0.101	0.246
(200,5,15)	0.1	0.001	0.001	0.005
	0.4	0.001	0.001	0.005
	0.8	0.018	0.031	0.153
(200,15,5)	0.1	0.366	0.351	0.395
	0.4	0.311	0.359	0.381
	0.8	0.265	0.432	0.461

Through numerical simulations we can see the following tendencies:

(i) As for Est and Approx, their precision deteriorates remarkably when $\rho = 0.8$.

(ii) The Est is bigger than the true value in all tables.

§4. Real Example

We apply our method to a real dataset of microarray data.

4.1. Leukemia dataset

Leukemia dataset used by Dudoit et al. (2002) contains gene expression level of 72 patients either suffering from acute lymphoblastic leukemia (47 cases) or acute myeloid leukemia (25 cases) and was obtained from Affymetrix oligonucleotide microarrays. Following the protocol in Dudoit et al. (2002), we preprocess the data by thresholding, filtering, a logarithmic transformation and standardization, so that the data finally comprise the expression $p = 3571$ genes. The dataset is publically available at

"http://www.broadinstitute.org/cgi-bin/cancer/datasets.cgi".

The normality assumption of the data set was checked the normality by QQ-plotting around 50 genes selected randomly in Srivastava and Kubokawa (2008). The results are nearly satisfactory.

4.2. Performance of ridge type discriminanation methods

In Dudoit et al. (2002), they use BW ratio criterion which is based on the ratio of the between-group to within-group sums of squares. For a gene j, $\mathrm{BW}(j) = b_{jj}/w_{jj}$, where $B = (N_1N_2/N)(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' = (b_{ij})$ and $W = \sum_{i=1}^{2}\sum_{j=1}^{N_i}(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)' = (w_{ij})$. Let K be the set of k indices with the largest BW ratios. In this paper, we choose $k = 500,\ 1000,\ 2000,\ 3000,\ 3571$. We investigate the EPMC of ridge type linear discriminant analysis:

$$\mathrm{RTLDA} : W_r = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'S_r^{-1}\{\boldsymbol{x} - \tfrac{1}{2}(\bar{\boldsymbol{x}}_1 + \bar{\boldsymbol{x}}_2)\} \lesseqgtr 0 \implies \boldsymbol{x} \in \Pi_1(\Pi_2).$$

From (2.4), we can estimate the EPMC of RTLDA as follows:

$$\hat{e}(2|1) = \Phi(\hat{\xi}).$$

Using the above estimator of EPMC and Leave-One-Out cross validation, we can check performance of RTLDA (Table 5).

Table 5. The estimator of EPMCs

k	$\hat{e}(1\|2)$	Leave-One-Out	$\hat{e}(2\|1)$	Leave-One-Out
500	0.008	0.080	0.008	0.042
1000	0.010	0.040	0.010	0.040
2000	0.011	0.040	0.011	0.040
3000	0.012	0.040	0.012	0.040
3571	0.012	0.040	0.012	0.040

§5. Conclusion

In this paper, we consider the classification problem for high-dimensional data. For high-dimensional data classification, due to the small number of observations and large number of dimension, classical LDA has sub-optimal performance corresponding to the singularity and instability of the pooled sample covariance matrix. Our modified LDA approach is RTLDA based on ridge type estimator of covariance matrix. Besides, we examined the performance of this discrimination method based on EPMC. In general, it is generally difficult to obtain an exact expression for the EPMC. Therefore, we consider an asymptotic approximation of EPMC under some assumptions about the parameter. By a results of the simulation, this approximation has good. In addition, the EPMC of RTLDA depends on the set $(\Delta_0, \Delta_1, a_{10}, a_{20})$ from our approximation of EPMC. We can say that the EPMC decreases if value of the ratio of $\Delta_0/\Delta_1^{1/2}$ becomes big as a rough guide. We understand that RTLDA shows the high performance from results on the real dataset. It was concluded that the RTLDA method can be used as effective classification tools in limited sample size and high-dimensional microarray classification problems.

Appendix

In this section, we prove Lemma 3 stated in Section 2. But before we begin these proofs, we state some preliminary results.

A.1. Preliminary results

Lemma A. 1. *Let A, B and D be $p \times p$ positive definite matrices, and let C be an $p \times p$ positive semi definite matrix. If $A = B - C$ and $\boldsymbol{a}$ is any $p \times 1$*

vector, then for $i \in N$, it holds that

$$\begin{array}{ll} \text{(i)} & \boldsymbol{a}'(DA)^i\boldsymbol{a} \le \boldsymbol{a}'(DB)^i\boldsymbol{a}. \\ \text{(ii)} & \operatorname{tr}(DA)^i \le \operatorname{tr}(DB)^i. \end{array}$$

Proof. Using Theorem 3.26 in Schott (1997), DA and DB are positive definite matrix and DC is positive semi definite matrix. Thus, we note that

$$\begin{aligned} \boldsymbol{a}'(DA)^i\boldsymbol{a} &= \boldsymbol{a}'DB(DA)^{i-1}\boldsymbol{a} - \boldsymbol{a}'DC(DA)^{i-1}\boldsymbol{a} \\ &\le \boldsymbol{a}'DB(DA)^{i-1}\boldsymbol{a} \\ &= \boldsymbol{a}'(DB)^2(DA)^{i-2}\boldsymbol{a} - \boldsymbol{a}'DBDC(DA)^{i-2}\boldsymbol{a} \\ &\le \boldsymbol{a}'(DB)^2(DA)^{i-2}\boldsymbol{a} \\ &\vdots \\ &= \boldsymbol{a}'(DB)^k(DA)^{i-k}\boldsymbol{a} - \boldsymbol{a}'(DB)^{k-1}DC(DA)^{i-k}\boldsymbol{a} \\ &\le \boldsymbol{a}'(DB)^k(DA)^{i-k}\boldsymbol{a} \\ &\vdots \\ &= \boldsymbol{a}'(DB)^i\boldsymbol{a} - \boldsymbol{a}'(DB)^{i-1}DC\boldsymbol{a} \\ &\le \boldsymbol{a}'(DB)^i\boldsymbol{a}. \end{aligned}$$

This proves (i) of Lemma A.1. It is noted that

$$\operatorname{tr}(DA)^i = \sum_{i=1}^{p} \boldsymbol{a}_i'(DA)^i\boldsymbol{a}_i,$$

where $\boldsymbol{a}_1 = (1,0,\ldots,0), \boldsymbol{a}_2 = (0,1,\ldots,0), \ldots, \boldsymbol{a}_p = (0,0,\ldots,1)$. Using (i) of Lemma A.1, we can easily check (ii). □

Lemma A. 2 (Srivastava (2005))**.** *Let $\hat{a}_1$ be as defined in Section 2. Then under the assumptions* A.1 *and* A.2*, asymptotically*

$$\sqrt{np}(\hat{a}_1 - a_{10}) \xrightarrow{d} N_1(0, 2a_{20}).$$

Here, the notation "$\xrightarrow{d}$" denotes convergence in distribution.

Proof. The proof is given in Srivastava (2005). □

Lemma A. 3. *Let $\hat{a}_1$ be as defined in Section 2. Then under assumptions* A.1 *and* A.2*, asymptotically*

$$\begin{array}{ll} \text{(i)} & \sqrt{np}(1/\hat{a}_1 - 1/a_{10}) \xrightarrow{d} N_1(0, 2a_{20}/a_{10}^4). \\ \text{(ii)} & \lim_{p\to\infty} 1/\hat{a}_1 = 1/a_{10} \textit{ in probability}. \end{array}$$

Proof. Using Lemma A.2 and the delta method, we can easily check (i). Using Continuous Mapping Theorem and (i) of Lemma 1, we can get (ii). This proves (ii) of Lemma A.3. □

A.2. Proof of Lemma 3

First, we show (i) of Lemma 3. $U/p^{\delta+1/2}$ can be expressed as

$$\text{(A. 1)}\quad \begin{aligned} U/p^{\delta+1/2} &= -\frac{1}{2p^{\delta+1/2}}\boldsymbol{\delta}'S_r^{-1}\boldsymbol{\delta} + \frac{1}{N^{\frac{1}{2}}p^{\delta+1/2}}\boldsymbol{\delta}'S_r^{-1}\boldsymbol{z}_1 \\ &\quad - \left(\frac{N_1}{NN_2p^{2\delta+1}}\right)^{\frac{1}{2}}\boldsymbol{\delta}'S_r^{-1}\boldsymbol{z}_2 + \frac{1}{(N_1N_2p^{2\delta+1})^{\frac{1}{2}}}\boldsymbol{z}_1'S_r^{-1}\boldsymbol{z}_2 \\ &\quad - \frac{N_1-N_2}{2N_1N_2p^{\delta+1/2}}\boldsymbol{z}_2'S_r^{-1}\boldsymbol{z}_2. \end{aligned}$$

We note that

$$\text{(A. 2)}\quad S_r^{-1} = n\{(\sqrt{p}\hat{a}_1)^{-1}I_p - (\sqrt{p}\hat{a}_1)^{-1}H_1(I_n + (\sqrt{p}\hat{a}_1)L^{-1})^{-1}H_1'\}.$$

Here, $nS = H_1'LH_1$, where $H_1'H_1 = I_n$ and $L = (\ell_1, \dots, \ell_n)$, an $n \times n$ diagonal matrix which contains the non-zero eigenvalues of nS. The first term of (A. 1) is expressed

$$\begin{aligned} \frac{1}{2p^{\delta+1/2}}\boldsymbol{\delta}'S_r^{-1}\boldsymbol{\delta} &= \frac{n}{2p^{\delta+1/2}}\boldsymbol{\delta}'\{(\sqrt{p}\hat{a}_1)^{-1}I_p \\ &\qquad - (\sqrt{p}\hat{a}_1)^{-1}H_1(I_n + (\sqrt{p}\hat{a}_1)L^{-1})^{-1}H_1'\}\boldsymbol{\delta}. \end{aligned}$$

Then we get from Lemmas 1 and 2,

$$\text{(A. 3)}\quad \frac{\boldsymbol{\delta}'S_r^{-1}\boldsymbol{\delta}}{2p^{\delta+1/2}} = \frac{n}{2p^{\delta}}\left(\frac{\boldsymbol{\delta}'\boldsymbol{\delta}}{pa_{10}}\right) + o_p(p^{-1/2}).$$

From Lemmas 1 and 2, we also note that

$$\text{(A. 4)}\quad \mathrm{E}\left[\frac{N_1-N_2}{2N_1N_2p^{\delta+1/2}}\boldsymbol{z}_2'S_r^{-1}\boldsymbol{z}_2\right] = \frac{(N_1-N_2)n}{2N_1N_2p^{\delta}} + o(p^{-1/2}).$$

Then, it is sufficient to show that

$$\text{(A. 5)}\quad \lim_{p\to\infty}\mathrm{E}\left[\left(\frac{N_1-N_2}{2N_1N_2p^{\delta+1/2}}\boldsymbol{z}_2'S_r^{-1}\boldsymbol{z}_2 - \frac{(N_1-N_2)n}{2N_1N_2p^{\delta}}\right)^2\right] = 0.$$

From Lemma A.1, (A.2) and the independency of $\boldsymbol{z}_2$ and $\hat{a}_1$, it can be expressed that

$$
\begin{aligned}
&\left(\frac{N_1-N_2}{2N_1N_2p^{\delta+1/2}}\right)^2 \mathrm{E}[(\boldsymbol{z}_2'S_r^{-1}\boldsymbol{z}_2-\sqrt{p}n)^2]\\
&=\left(\frac{N_1-N_2}{2N_1N_2p^{\delta+1/2}}\right)^2 \mathrm{E}[(\mathrm{tr}(\Sigma S_r^{-1}))^2+2\,\mathrm{tr}(\Sigma S_r^{-1}\Sigma S_r^{-1})\\
&\qquad\qquad -2\sqrt{p}n\,\mathrm{tr}(\Sigma S_r^{-1})+pn^2]\\
&\le\left(\frac{N_1-N_2}{2N_1N_2p^{\delta+1/2}}\right)^2 \mathrm{E}\left[\left(\frac{\sqrt{p}na_1}{\hat{a}_1}\right)^2+\frac{2n^2a_2}{\hat{a}_1^2}\right.\\
&\quad\left.-2\left(\frac{pn^2a_1}{\hat{a}_1}-\frac{n^2\,\mathrm{tr}((I_n+(\sqrt{p}\hat{a}_1)L^{-1})H_1'\Sigma H_1)}{\hat{a}_1}\right)+pn^2\right].
\end{aligned}
$$

From Lemmas 1, 2 and A.3, we can evaluate

$$
\begin{aligned}
&\left(\frac{N_1-N_2}{2N_1N_2p^{\delta+1/2}}\right)^2 \mathrm{E}\left[\left(\frac{\sqrt{p}na_1}{\hat{a}_1}\right)^2+\frac{2n^2a_2}{\hat{a}_1^2}\right.\\
&\quad\left.-2\left(\frac{pn^2a_1}{\hat{a}_1}-\frac{n^2\,\mathrm{tr}((I_n+(\sqrt{p}\hat{a}_1)L^{-1})H_1'\Sigma H_1)}{\hat{a}_1}\right)+pn^2\right]\\
&=\left(\frac{N_1-N_2}{2N_1N_2p^{\delta+1/2}}\right)^2 \mathrm{E}\left[(\sqrt{p}n)^2+\frac{2n^2a_2}{a_1^2}\right.\\
&\qquad\qquad\left.-2\left(pn^2-\frac{n^3a_2}{(1+1/\sqrt{p})a_1^2}\right)+pn^2\right],
\end{aligned}
$$

as $p\to\infty$. Therefore,

$$
\begin{aligned}
&\left(\frac{N_1-N_2}{2N_1N_2p^{\delta+1/2}}\right)^2 \lim_{p\to\infty}\mathrm{E}[(\boldsymbol{z}_2'S_r^{-1}\boldsymbol{z}_2-\sqrt{p}n)^2]\\
&\le\left(\frac{N_1-N_2}{2N_1N_2p^{\delta+1/2}}\right)^2\left(\frac{2n^2a_{20}}{a_{10}^2}-\frac{2n^3a_2}{(1+1/\sqrt{p})a_1^2}\right)\\
&=O(p^{-\delta-1}).
\end{aligned}
$$

This proves (A.5). Using (A.4), (A.5) and Marcov's inequality

$$
\begin{aligned}
&\Pr\left\{\left|\frac{N_1-N_2}{2N_1N_2p^{\delta+1/2}}\boldsymbol{z}_2'S_r^{-1}\boldsymbol{z}_2-\frac{(N_1-N_2)n}{2N_1N_2p^{\delta}}\right|>\varepsilon\right\}\\
&\le\frac{\{(N_1-N_2)/(2N_1N_2p^{\delta+1/2})\}^2\,\mathrm{E}[(\boldsymbol{z}_2'S_r^{-1}\boldsymbol{z}_2-\sqrt{p}n)^2]}{\varepsilon^2}\\
&=0 \text{ as } p\to\infty.
\end{aligned}
$$

It follows that

$$
\text{(A. 6)}\qquad \frac{N_1-N_2}{2N_1N_2p^{\delta+1/2}}\boldsymbol{z}_2'S_r^{-1}\boldsymbol{z}_2=\frac{(N_1-N_2)n}{2N_1N_2p^{\delta}}+o_p(p^{-1/2}).
$$

With the similar evaluation method of the last term of (A.1), second term, third term and forth term of the (A.1) are

$$\text{(A. 7)} \qquad \frac{1}{(Np^{2\delta+1})^{\frac{1}{2}}}\boldsymbol{\delta}' S_r^{-1}\boldsymbol{z}_1 = o_p(p^{-1/2}).$$

$$\text{(A. 8)} \qquad \left(\frac{N_1}{NN_2p^{2\delta+2}}\right)^{\frac{1}{2}}\boldsymbol{\delta}' S_r^{-1}\boldsymbol{z}_2 = o_p(p^{-1/2}).$$

$$\text{(A. 9)} \qquad \frac{1}{(N_1N_2p^{2\delta+1})^{\frac{1}{2}}}\boldsymbol{z}_1' S_r^{-1}\boldsymbol{z}_2 = o_p(p^{-1/2}).$$

Combining (A.3) and (A.6)-(A.9), it holds that

$$U/p^{\delta+1/2} = -\frac{n}{2p^{\delta}}\left(\frac{\boldsymbol{\delta}'\boldsymbol{\delta}}{pa_{10}} + \frac{N_1-N_2}{N_1N_2}\right) + o_p(p^{-1/2}).$$

This proves (i) of Lemma 3.

Next, we show (ii) of Lemma 3. $V/p^{2\delta}$ can be expressed as

$$\text{(A. 10)} \quad V/p^{2\delta} = \frac{1}{p^{2\delta}}\boldsymbol{\delta}' S_r^{-1}\Sigma S_r^{-1}\boldsymbol{\delta} + 2\left(\frac{N}{N_1N_2p^{4\delta}}\right)^{\frac{1}{2}}\boldsymbol{\delta}' S_r^{-1}\Sigma S_r^{-1}\boldsymbol{z}_2 + \frac{N}{N_1N_2p^{2\delta}}\boldsymbol{z}_2' S_r^{-1}\Sigma S_r^{-1}\boldsymbol{z}_2.$$

From Lemmas 1 and 2, the first term of (A.10) is evaluated as follows:

$$\text{(A. 11)} \qquad \frac{1}{p^{2\delta}}\boldsymbol{\delta}' S_r^{-1}\Sigma S_r^{-1}\boldsymbol{\delta} = \frac{n^2(\boldsymbol{\delta}'\Sigma\boldsymbol{\delta}/p)}{p^{2\delta}a_{10}^2} + o_p(p^{-1/2}).$$

From Lemmas 1 and 2, we also note that

$$\text{(A. 12)} \quad \mathrm{E}\left[\frac{N}{N_1N_2p^{2\delta}}\boldsymbol{z}_2' S_r^{-1}\Sigma S_r^{-1}\boldsymbol{z}_2\right] = \frac{Nn^2a_{20}}{N_1N_2p^{2\delta}a_{10}^2} + o(p^{-1/2}).$$

Then, it is sufficient to show that

$$\text{(A. 13)} \quad \lim_{p\to\infty}\mathrm{E}\left[\left(\frac{N}{N_1N_2p^{2\delta}}\boldsymbol{z}_2' S_r^{-1}\Sigma S_r^{-1}\boldsymbol{z}_2 - \frac{Nn^2a_{20}}{N_1N_2p^{2\delta}a_{10}^2}\right)^2\right] = 0.$$

From Lemma A.1, (A.2) and the independency of $\boldsymbol{z}_2$ and $\hat{a}_1$, it can be expressed that

$$\begin{aligned}
&\left(\frac{N}{N_1N_2p^{2\delta}}\right)^2 \mathrm{E}\left[\left(\boldsymbol{z}_2' S_r^{-1}\Sigma S_r^{-1}\boldsymbol{z}_2 - \frac{n^2 a_{20}}{a_{10}^2}\right)^2\right] \\
&= \left(\frac{N}{N_1N_2p^{2\delta}}\right)^2 \mathrm{E}\Big[(\mathrm{tr}(\Sigma S_r^{-1})^2)^2 + 2\,\mathrm{tr}(\Sigma S_r^{-1})^4 \\
&\qquad\qquad - \frac{2n^2a_{20}\,\mathrm{tr}(\Sigma S_r^{-1})^2}{a_{10}^2} - \left(\frac{n^2 a_2}{a_{10}^2}\right)^2\Big] \\
&\le \left(\frac{N}{N_1N_2p^{2\delta}}\right)^2 \mathrm{E}\Big[\frac{n^4a_2^2}{\hat{a}_1^4} + \frac{2n^4a_4}{p\hat{a}_1^4} - \frac{2n^2a_{20}}{a_{10}^2}\Big(\frac{n^2a_{20}}{\hat{a}_1^2} \\
&\qquad\qquad - \frac{2n^2\,\mathrm{tr}((I_n + (\sqrt{p}\hat{a}_1)L^{-1})^{-1}H_1'\Sigma^2H_1))}{p\hat{a}_1^2} \\
&\qquad\qquad + \frac{n^2\,\mathrm{tr}(\{(I_n + (\sqrt{p}\hat{a}_1)L^{-1})^{-1}H_1'\Sigma H_1\}^2)}{p\hat{a}_1^2}\Big) \\
&\qquad\qquad - \frac{n^4a_{20}^2}{a_{10}^4}\Big]\ (\equiv C).
\end{aligned}$$

From Lemmas 1, 2 and A.3, we can evaluate

$$\begin{aligned}
C = \mathrm{E}\Bigg[&\left(\frac{N}{N_1N_2p^{2\delta}}\right)^2\Bigg(\frac{n^4a_{20}^2}{a_{10}^4} + \frac{2n^4a_{40}}{pa_{10}^4} - \frac{2n^4a_{20}^2}{a_{10}^4} \\
&+ \frac{4n^4a_{20}a_{30}}{(1+1/\sqrt{p})pa_{10}^5} - \frac{2n^4a_{20}^3}{(1+1/\sqrt{p})pa_{10}^6} + \frac{n^4a_{20}^2}{a_{10}^4}\Bigg)\Bigg]
\end{aligned}$$

as $p \to \infty$. Therefore,

$$\begin{aligned}
&\left(\frac{N}{N_1N_2p^{2\delta}}\right)^2 \lim_{p\to\infty} \mathrm{E}\left[\left(\boldsymbol{z}_2' S_r^{-1}\Sigma S_r^{-1}\boldsymbol{z}_2 - \frac{n^2 a_{20}}{a_{10}^2}\right)^2\right] \\
&\le \left(\frac{N}{N_1N_2p^{2\delta}}\right)^2\left(\frac{2n^4a_{40}}{pa_{10}^4} + \frac{4n^4a_{20}a_{30}}{(1+1/\sqrt{p})pa_{10}^5} - \frac{2n^4a_{20}^3}{(1+1/\sqrt{p})pa_{10}^6}\right) \\
&= O(p^{-1-2\delta}).
\end{aligned}$$

This proves (A.13). Using (A.12), (A.13) and Marcov's inequality

$$\begin{aligned}
&\Pr\left\{\left|\frac{N}{N_1N_2p^{2\delta}}\boldsymbol{z}_2' S_r^{-1}\Sigma S_r^{-1}\boldsymbol{z}_2 - \frac{Nna_{20}}{N_1N_2p^{2\delta}a_{10}^2}\right| > \varepsilon\right\} \\
&\le \frac{\mathrm{E}[\{N/(N_1N_2p^{2\delta})\boldsymbol{z}_2' S_r^{-1}\Sigma S_r^{-1}\boldsymbol{z}_2 - (Nn^2a_{20})/(N_1N_2p^{2\delta}a_{10}^2)\}^2]}{\varepsilon^2} \\
&= 0 \text{ as } p \to \infty.
\end{aligned}$$

Hence, it follows that

$$\text{(A. 14)} \qquad \frac{N}{N_1N_2p^{2\delta}}\boldsymbol{z}_2'S_r^{-1}\boldsymbol{z}_2 = \frac{Nn^2a_{20}}{N_1N_2p^{2\delta}a_{10}^2} + o_p(p^{-1/2}).$$

With the similar evaluation method of the last term of (A.10), second term of (A.10) is

$$\text{(A. 15)} \qquad \left(\frac{N}{N_1N_2p^{4\delta}}\right)^{\frac{1}{2}}\boldsymbol{\delta}'S_r^{-1}\Sigma S_r^{-1}\boldsymbol{z}_2 = o_p(p^{-1/2}).$$

Combining (A.11), (A.14) and (A.15), it holds that

$$V/p^{2\delta} = \frac{n^2}{p^{2\delta}}\left(\frac{\boldsymbol{\delta}'\Sigma\boldsymbol{\delta}}{pa_{10}^2} + \frac{Na_{20}}{N_1N_2a_{10}^2}\right) + o_p(p^{-1/2}).$$

This proves (ii) of Lemma 3. □

Acknowledgements

I would like to thank the referee for suitable comments and careful reading. In addition, I am greatful to Professor Takashi Seo for his advice and encouragement.

References

[1] Dudoit, S., Fridlyand, J. and Speed, T.P. (2002). Comparison of discrimination methods for the classification of tumors using gene expression data. *J. Amer. Statist. Assoc.*, **97**, 77-87.

[2] Fujikoshi, Y. and Seo, T. (1998). Asymptotic approximations for EPMC's of the linear and the quadratic discriminant function when the sample size and the dimension are large. *Random Oper. Stoch. Equ.*, **6**, 269-280.

[3] Fujikoshi, Y. (2000). Error bounds for asymptotic approximations of the linear discriminant function when the sample size and dimensionality are large. *J. Multivariate Anal.*, **73**, 1-17.

[4] Lachenbruch, P. A. (1968). On Expected Probabilities of Misclassification in Discriminant Analysis, Necessary Sample Size, and a Relation with the Multiple Correlation Coefficient. *Biometrics*, **24**, 823-834.

[5] Raudys, S. (1972). On the amount of a priori information in construction of a classification algorithm. *Engrg. Cybernetics*, **4**, 711-718.

[6] Schott, J. R. (1997). *Matrix analysis for statistics.* Wiley Series in Probability and Statistics.

[7] Siotani, M. (1982). Large sample approximations and asymptotic expansions of classification statistic. *Handbook of Statistics* **2** (P. R. Krishnaiah and L. N. Kanal, Eds.), North-Holland Publishing Company, 61-100.

[8] Srivastava, M. S. (2005). Some tests concerning the covariance matrix in high dimensional data. *J. Japan Statist. Soc.*, **35**, 251-272.

[9] Srivastava, M. S. (2007). Multivariate theory for analyzing high dimensional data. *J. Japan Statist. Soc.*, **37**, 53-86.

[10] Srivastava, M. S. and Kubokawa, T. (2007). Comparison of discrimination methods for high dimensional data. *J. Japan Statist. Soc.*, **37**, 123-134.

[11] Srivastava, M. S. and Kubokawa, T. (2008). Akaike information criterion for selecting components of the mean vector in high dimensional data with fewer observations. *J. Japan Statist. Soc.*, **38**, 259-283.

Masashi Hyodo
Graduate School of Science, Tokyo University of Science
1-3 Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan

E-mail: j1108702@ed.kagu.tus.ac.jp